# *Organization of Genome Data into Pathways and Networks*

**Peter D. Karp, Ph.D.**

**Bioinformatics Research Group**

**SRI International**

**pkarp@ai.sri.com**

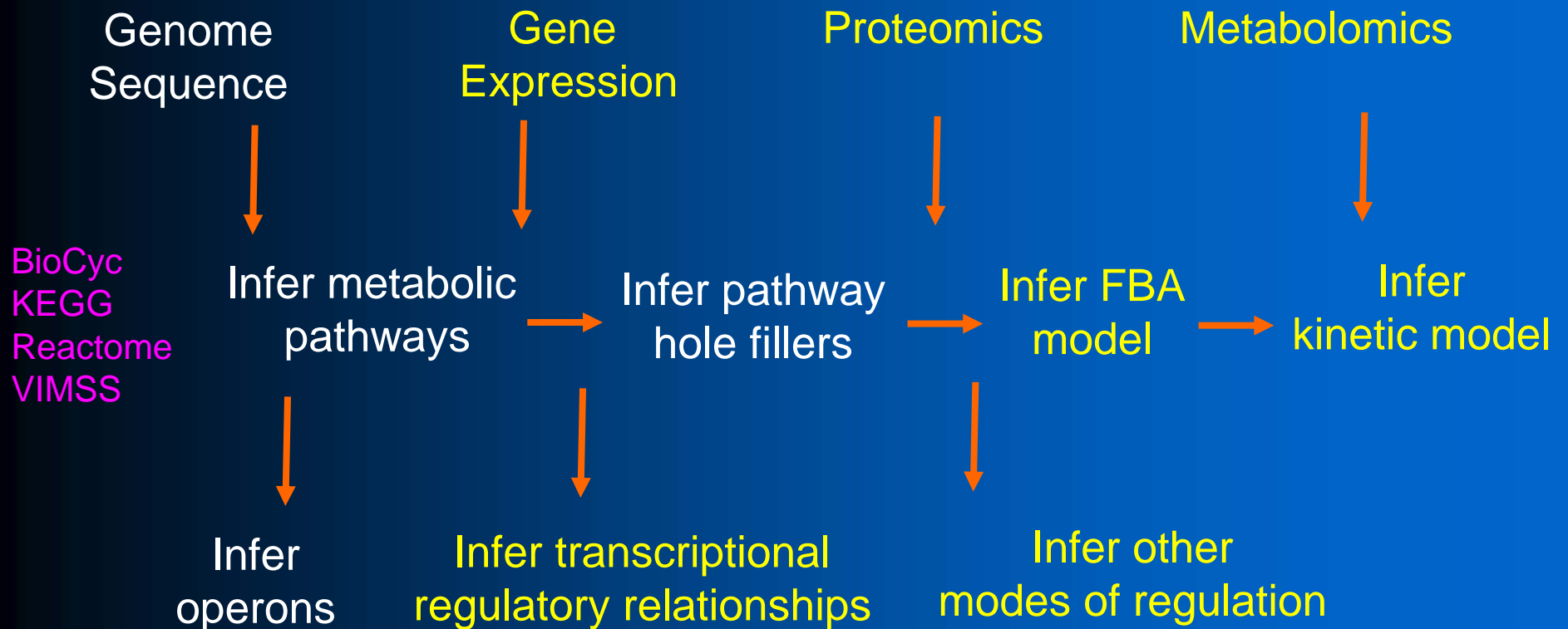# *Organization of Genome Data into Pathways and Networks*

- Summarize state of the art, existing approaches

- Opportunities for new research directions

- Limitations in current approaches

*Assigning Genes to Pathways*

SRI International
Bioinformatics

Genome
Sequence

Gene
Expression

Proteomics

Metabolomics

BioCyc
KEGG
Reactome
VIMSS

Infer metabolic
pathways

Infer pathway
hole fillers

Infer FBA
model

Infer
kinetic model

Infer
operons

Infer transcriptional
regulatory relationships

Infer other
modes of regulation

# *Limitations in Pathway Assignment*

- **Inference of metabolic pathways**
  - Quality of genome annotations
    - False positives
    - False negatives (ORFs and missing multiple functions)
    - Lack of controlled vocabulary in many genome annotations
    - Lack of probability values in genome annotations
  - Many enzymes within pathways can never be present in a genome annotation – never sequenced

# *Experimental/Computational Partnership To Improve Genome Annotations*

- **Focused effort proposed to**
  - Experimentally verify computational predictions of functions for genes of unknown function
  - Seek which genes encode functions with no associated sequence
  - Capture computational and experimental results in common database
  -
- **Roberts, R.J., Karp, P.D., Kasif, S., Linn, S., and Buckley, M.R. ``An Experimental Approach to Genome Annotation,'' (2004) published by the American Society for Microbiology, http://www.asm.org/academy/index.asp?bid=32664.**

- **Roberts, R., "Identifying protein function - A call for community action," PLoS Biology 2:E42 2004 http://biology.plosjournals.org/plosonline/?request=get-document&doi=10.1371/journal.pbio.0020042**

- **Karp, P.D., "Call for an enzyme genomics initiative" Genome Biology 5:401.1-3 http://genomebiology.com/2004/5/8/401**

# *Limitations in Pathway Assignment*

- **Inference of metabolic pathways**
    - Prediction of novel pathways
    - Pathway databases don't yet contain all experimentally elucidated pathways
    - Choosing among multiple pathway variants
    - Lack of experimental testing of predicted pathways; results would likely lead to improvements in prediction algorithms

# *Curation of Organism-Specific Pathway Models*

- Centralized in a single group?

- Distributed across many groups?

- Automated mining of pathways from the literature